mRNA diffusion explains protein gradients in *Drosophila* early development

Rui Dilão[1,*], Daniele Muraro[1]

1 Nonlinear Dynamics Group, Instituto Superior Técnico Av. Rovisco Pais, 1049-001 Lisbon, Portugal

∗ **E-mail: Corresponding rui@sd.ist.utl.pt**

Abstract

We propose a new model describing the production and the establishment of the stable gradient of the Bicoid protein along the antero-posterior axis of the embryo of *Drosophila*. In this model, we consider that *bicoid* mRNA diffuses along the antero-posterior axis of the embryo and the protein is produced in the ribosomes localized near the syncytial nuclei. Bicoid protein stays localized near the syncytial nuclei as observed in experiments. We calibrate the parameters of the mathematical model with experimental data taken during the cleavage stages 11 to 14 of the developing embryo of *Drosophila*. We obtain good agreement between the experimental and the model gradients, with relative errors in the range $5-8\%$. The inferred diffusion coefficient of *bicoid* mRNA is in the range $4.6\times 10^{-12}-1.5\times 10^{-11}$ $\mathrm{m^2 s^{-1}}$, in agreement with the theoretical predictions and experimental measurements for the diffusion of macromolecules in the cytoplasm. We show that the model based on the mRNA diffusion hypothesis is consistent with the known observational data, supporting the recent experimental findings of the gradient of *bicoid* mRNA in *Drosophila* [Spirov *et al.* (2009) *Development* 136:605-614].

Introduction

In *Drosophila* early development, *bicoid* mRNA of maternal origin is deposited in one of the poles of the egg, determining the anterior tip of the embryo, (Frigerio *et al.*, 1986; Driever and Nüsslein-Volhard, 1988). The deposition of the mRNA is done during oogenesis and is transported into the oocyte along microtubules, (Saxton, 2001). After fertilization and deposition of the egg, and during the first 14 nuclear divisions of the developing embryo, *bicoid* mRNA of maternal origin is translated into protein in the ribosomes.

During the interphases following the 11th nuclear division up to the 14th, the concentration of Bicoid protein distributes non-uniformly along the antero-posterior axis of the syncytial blastoderm. Bicoid has higher concentration near the anterior pole of the embryo, and its local concentration decreases as the distance to the anterior pole increases. This is called the Bicoid protein gradient, (Nüsslein-Volhard, 1992).

As, during oogenesis, *bicoid* mRNA is deposited near the anterior pole of the embryo, it is implicitly assumed that Bicoid protein is produced in the ribosomes of the

nuclei localized near the anterior pole of the embryo, and then diffuses through the syncytial blastoderm. Eventually, this protein diffusion could be facilitated by the absence of cellular membranes in the syncytial phase of the developing embryo. Driever and Nüsslein-Volhard (Driever and Nüsslein-Volhard, 1988) argued that the protein gradient is generated by protein diffusion and degradation throughout the embryo. Later, Nüsslein-Volhard (Nüsslein-Volhard, 1992, 2006) emphasized that the Bicoid protein diffuses away from the site of its production, the local mRNA deposition region. The theoretical possibility of this mechanism of morphogenesis goes back to the work of Alan Turing in the fifties (Turing, 1952), and has been further discussed and analyzed by Wolpert (Wolpert, 1969), Crick (Crick, 1970) and Meinhardt (Meinhardt, 1977). Experimental measurements within mammalian cells (Wojcieszyn *et al.*, 1981; Mastro *et al.*, 1984) and theoretical analysis (Crick, 1970) suggested that diffusion coefficients of macromolecules in the cytoplasm are in the range $10^{-11} - 10^{-13}$ $\mathrm{m^2 s^{-1}}$.

However, there are several open questions related with the establishment of the stable gradient of protein Bicoid along the antero-posterior axis of the embryo of *Drosophila*. Experimental observations during cleavage stages $11 - 14$, and before cellularization that occurs at the end of cleavage stage 14, show that Bicoid protein is always localized around the syncytial nuclei, Figure 1. This can be seen in the embryo data sets b18, ab17, ab16, ab12, ab14, ab9, ad13, ab8 of the FlyEx database (Kozlov *et al.*, 2000; Myasnikova *et al.*, 1999, 2001; Poustelnikova *et al.*, 2004; Pisarev *et al.*, 2009, http://flyex.ams.sunysb.edu/flyex/). This fact suggests that dispersal effects driven by molecular collisions with Bicoid protein — Brownian motion — do not play a significant role in the establishment of the gradient of Bicoid. As Brownian motion is the driven mechanism of diffusion dispersal, (Murray, 1993, chap. 9), it is difficult to understand how the diffusion of a protein produces a strongly localized protein concentration around the syncytial nuclei and during successive mitotic cycles. On the other hand, as Bicoid protein is produced and attains a steady state during the first cleavage cycles, its localization near the nuclear envelops suggest that ribosomes are also localized near the nucleus. If ribosomes were not localized near the nuclear envelopes of the syncytial nucleus, protein in the inter-nuclear regions of the embryo would be observed.

As argued by Kerszberg and Wolpert (Kerszberg and Wolpert, 2007), there is not a clear experimental evidence of protein degradation, a necessary mechanism for the establishment of a steady protein gradient in models based on protein diffusion.

Houchmandzadeh *et al.* (Houchmandzadeh *et al.*, 2005), reported the constant Bicoid protein concentration during cleavage cycles 12-14, suggesting the stability of protein concentration during an important developmental period. On the other hand, the protein diffusion hypothesis lead to some quantitative contradictory facts. For example, in recent experiments, the hypothetical inferred cytoplasmic diffusion coefficient of the Bicoid protein during the cleavage stage 13 is of the order of 0.3×10^{-12} $\mathrm{m^2 s^{-1}}$, (Gregor *et al.*, 2007). However, during the first cleavage stages of the developing embryo, the Bicoid protein reaches a steady state in 90 minutes (end of cleavage stage 9), (Gregor *et al.*, 2007),

and a simple estimate with the Houchmandzadeh *et al.* (Houchmandzadeh *et al.*, 2005) model shows that the diffusion coefficients must be of the order of 2×10^{-12} $m^2 s^{-1}$, (Gregor *et al.*, 2007). This value for the diffusion coefficient is one order of magnitude larger than the value inferred from experiments. This discrepancy between model estimates and observation needs a clear explanation, (Reinitz, 2007).

Here, with a mathematical model, we show that the observed gradient of the Bicoid protein can be explained by the diffusion of *bicoid* mRNA, and Bicoid protein stays localized near the nuclei of the syncytial blastoderm of the embryo of *Drosophila*. This explains the absence or the very low level of Bicoid concentration in the regions between the nuclei during the first stage of development of *Drosophila*. We determine a scaling relation between the mRNA diffusion coefficient, the embryo length and the mRNA degradation rate, enabling the precise determination of the diffusion coefficient of *bicoid* mRNA. In this model, it is not necessary to introduce the morphogen degradation hypothesis for protein, and the steady gradient of protein is reached after the complete translation of mRNA of maternal origin.

The mRNA localization mechanism in the embryo of *Drosophila* has been analyzed experimentally by several authors, and (Saxton, 2001) argues that the relative small size of mRNA suggests that random diffusion and specific anchoring to the cytoskeleton in a target area might suffice for localization in the syncytial blastoderm. Cha *et al.* (Cha *et al.*, 2001) reported rapid saltatory movements in injected *bicoid* mRNA in the embryo, followed by dispersion without localization. Other effects of diffusing mRNA has been reported by Forrest and Gavis (Forrest and Gavis, 2003) for the *nanos* mRNA. More recently, Spirov *et al.* (Spirov *et al.*, 2009) have shown that a *bicoid* mRNA gradient exists along the antero-posterior axis of the embryo of *Drosophila*, completely changing our current views of this *Drosophila* developmental pathway. The model presented here corroborates these experimental facts, is consistent with the experimental facts and observations, and fits the experimental data with high accuracy.

Results

We now derive a mRNA diffusion model and we show that experimental protein gradients are well fitted in this framework. This shows that a mechanism of mRNA mobility (diffusion) is enough to explain protein gradients.

A mRNA diffusion model

It is an experimental fact that mRNA of maternal origin is deposited in a small region of the embryo of *Drosophila*, defining the anterior pole of the fertilized egg, (Driever and Nüsslein-Volhard, 1988). After the deposition of *bicoid* mRNA, we assume that *bicoid* mRNA of maternal origin disperses within the embryo and this process is simultaneous

with the successive cleavage stages. Then, the protein is produced in the ribosomes that are near the nuclear membranes of the nuclei in the syncytium. Representing by $R(x,t)$ the concentration of *bicoid* mRNA along the one-dimensional antero-posterior axis (x) of the embryo, and by $B(x,t)$ the concentration of Bicoid protein, the equations describing the production of Bicoid from mRNA are,

$$\begin{cases} \dfrac{\partial R}{\partial t} &= -dR + D\dfrac{\partial^2 R}{\partial x^2} \\ \dfrac{\partial B}{\partial t} &= aR \end{cases} \tag{1}$$

where a is the rate of production of Bicoid from mRNA, d is the degradation rate of *bicoid* mRNA, and D is the diffusion coefficient of *bicoid* mRNA in the cytoplasm. Using the mass action law, this simple model is straightforwardly derived from the rate mechanisms,

$$\begin{aligned} R &\xrightarrow{a} B + R \\ R &\xrightarrow{d} \end{aligned} \tag{2}$$

and then the diffusion term is added to the mRNA rate equation. In general, one molecule of mRNA can produce more than one molecule of protein, implying that $d < a$. If one molecule of mRNA produces one molecule of protein then, in the mean, we have $d = a$.

We consider that the length of the antero-posterior axis of the embryo is L, and so $x \in [0, L]$. We take zero flux boundary conditions, $\frac{\partial R}{\partial x}(x = 0, t) = \frac{\partial R}{\partial x}(x = L, t) = 0$, and $\frac{\partial B}{\partial x}(x = 0, t) = \frac{\partial B}{\partial x}(x = L, t) = 0$, for every $t \geq 0$. The protein initial condition is $B(x, t = 0) = 0$, and the initial distribution of mRNA is,

$$R(x, t = 0) = \begin{cases} A > 0 & \text{if} \quad 0 \leq \ell_1 \leq x \leq \ell_2 \leq L \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where A, ℓ_1 and ℓ_2 are constants. The function $R(x, t = 0)$ describes the initial distribution of *bicoid* mRNA of maternal origin deposited in the region of the embryo $[\ell_1, \ell_2] \subset [0, L]$. The concentration of mRNA of maternal origin deposited in the embryo is then $A(\ell_2 - \ell_1)$. In this model, *bicoid* mRNA has a fixed initial concentration, and the Bicoid protein does not degrade.

Equation (1) with the initial condition (3), and the zero flux boundary conditions define the mRNA diffusion model. This model is linear, and has solutions that can be determined explicitly. Now, we will show that, within this simple model, Bicoid protein attains a gradient like steady state along the embryo.

By standard Fourier analysis techniques, see for example (Alves and Dilão, 2006) or

(Haberman, 1998), the solution of the first equation in (1) is,

$$\begin{aligned} R(x,t) &= A\frac{\ell_2-\ell_1}{L}e^{-dt} \\ &+2A\sum_{i=1}^{\infty}\frac{e^{-dt-\frac{n^2\pi^2}{L^2}Dt}}{n\pi}\cos\left(\frac{n\pi x}{L}\right)\left(\sin\left(\frac{n\pi\ell_2}{L}\right)-\sin\left(\frac{n\pi\ell_1}{L}\right)\right) \end{aligned} \tag{4}$$

The solution of the second equation in (1) is,

$$B(x,t) = B(x,t=0) + a\int_0^t R(x,s)\,ds \tag{5}$$

In the limit $t\to\infty$, the equilibrium or steady solution of the Bicoid protein is calculated from (4) and (5), and we obtain,

$$\begin{aligned} B_{eq}(x) &= a_1\frac{\ell_2-\ell_1}{L} \\ &+2a_1\sum_{i=1}^{\infty}\frac{1}{n\pi+\frac{n^3\pi^3}{a_2^2}}\cos\left(\frac{n\pi x}{L}\right)\left(\sin\left(\frac{n\pi\ell_2}{L}\right)-\sin\left(\frac{n\pi\ell_1}{L}\right)\right) \end{aligned} \tag{6}$$

where,

$$a_1 = A\frac{a}{d}\,,\ a_2^2 = d\frac{L^2}{D} \tag{7}$$

and we have introduced into (5) the protein initial condition $B(x,t=0)=0$. A simple calculation shows that the solution (6) can be written as, (Alves and Dilão, 2006),

$$\begin{aligned} B_{eq}(x) &= 2\frac{a_1}{e^{2a_2/L}-1}\cosh(a_2\frac{x}{L})\left(\sinh(a_2\frac{\ell_2}{L})-\sinh(a_2\frac{\ell_1}{L})\right) \\ &+\frac{a_1}{2}\left(e^{-a_2(x+\ell_1)/L}-e^{-a_2(x+\ell_2)/L}\right)+I(x) \end{aligned} \tag{8}$$

where,

$$I(x) = \begin{cases} a_1\left(e^{-a_2(\ell_1-x)/L}-e^{-a_2(\ell_2-x)/L}\right)/2, & \text{if } x<\ell_1 \\ a_1-\frac{a_1}{2}\left(e^{-a_2(x-\ell_1)/L}+e^{-a_2(\ell_2-x)/L}\right), & \text{if } \ell_1\le x\le\ell_1 \\ a_1\left(e^{-a_2(x-\ell_2)/L}-e^{-a_2(x-\ell_1)/L}\right)/2, & \text{if } x>\ell_2 \end{cases} \tag{9}$$

The steady solution (8)-(9) describes the gradient of the Bicoid protein along the antero-posterior axis of the embryo of *Drosophila*. This solution depends on the set of five parameters a_1, a_2, L, ℓ_1 and ℓ_2, to be calibrated with experimental data. The constants a_1 and a_2^2 given by (7) define scaling relations of the embryo.

In order to compare the model predictions with the experimental data, the next step is to calibrate the parameters of the mRNA diffusion model (8)-(9) with the available experimental data for the gradient of the Bicoid protein.

Calibration of the mRNA diffusion model with the experimental data

To calibrate the parameters of the mRNA diffusion model (8)-(9) with the experimental data, we use the data available in the FlyEx database, (Poustelnikova *et al.*, 2004; Pisarev *et al.*, 2009). We considered Bicoid gradients for cleavage stages 11-14, and we make the additional assumption that during these cleavage stages, the concentration of the Bicoid protein is in the steady state, (Houchmandzadeh *et al.*, 2005).

We have fitted the data sets of the FlyEx database with the equilibrium distribution of Bicoid protein given by (8)-(9). The fitted functions are represented in Figure 2. In this figure, we show the concentration of Bicoid protein along the antero-posterior axis of *Drosophila*, for several embryos and in consecutive developmental stages. Due to the particular form of model prediction (8)-(9), the fitted parameter values are $\vec{\alpha} = (a_1, a_2, \ell_1/L, \ell_2/L)$. The parameter values of the different data sets are shown in Table 1, and correspond to the global minima of the fitness function $\chi^2(\vec{\alpha})$, introduced below in (15).

Table 1: **Fitted model parameters for the protein Bicoid antero-posterior distributions**

		a_1	a_2	ℓ_1/L	ℓ_2/L	$\sqrt{\chi^2_m/B^2_{max}}$	n	d (s^{-1})	$Aa(\ell_2-\ell_1)/L$
a)	ab18 (11)	345.2	4.69	0.03	0.20	0.06	30	8.8×10^{-4}	5.2×10^{-2}
b)	ab17 (12)	894.4	4.50	0.06	0.14	0.06	70	8.1×10^{-4}	5.8×10^{-2}
c)	ab16 (13)	684.2	5.51	0.06	0.15	0.08	152	1.2×10^{-3}	7.4×10^{-2}
d)	ab12 (14-1)	927.6	4.82	0.06	0.14	0.08	309	9.2×10^{-4}	6.8×10^{-2}
e)	ab14 (14-2)	3414.7	4.38	0.05	0.07	0.08	314	7.7×10^{-4}	5.3×10^{-2}
f)	ab9 (14-3)	1191.6	4.39	0.05	0.10	0.07	343	7.7×10^{-4}	4.6×10^{-2}
g)	ad13 (14-4)	470.4	3.02	0.06	0.19	0.05	324	3.6×10^{-4}	2.2×10^{-2}
h)	ab8 (14-5)	3271.7	4.25	0.08	0.09	0.07	332	7.2×10^{-4}	2.4×10^{-2}

Parameters values that best fit the experimental distribution of Bicoid protein shown in Figure 2 with the equilibrium distribution (8)-(9). In the first column, we show the data sets and the corresponding cleavage stages. The parameters a_1, a_2, ℓ_1/L and ℓ_2/L have been determined with the swarm algorithm described in the Materials and Methods section. The lengths of the embryos have been rescaled to the value $L = 1$. $\sqrt{\chi^2_m/B^2_{max}}$ is an estimate of the relative error of the fits, and n is the number of data points in the corresponding graphs in Figure 2. The parameter d has been determined by (7) with the estimated diffusion coefficient $D = 10^{-11}$ m^2s^{-1} of *bicoid* mRNA and $L = 0.5\times10^{-3}$ m. The parameter $Aa(\ell_2-\ell_1)/L$ has been determined with (10).

The quality of the fits of Figure 2 has been evaluated with the fitness function (15). Denoting by B_{max} the maximum value of each experimental data set, the mean relative

error of a fit is estimated by the quantity, $\sqrt{\chi^2_m/B^2_{max}}$, where $\chi^2_m = \min_{\vec{\alpha}\in S}\chi^2(\vec{\alpha})$. In Table 1, we show the mean relative errors of the fits, and the number of points (n) in each data set. The mean relative errors between the theoretical predictions (8)-(9) and the experimental data sets of Figure 2 are in the range $5\% - 8\%$, showing a remarkable agreement between the model prediction and the experimental data.

To determine the values of the parameter d in Table 1, we have fixed the embryo length to the value $L = 0.5 \times 10^{-3}$ m, (Nüsslein-Volhard, 2006, cap. iv). For the diffusion coefficient of *bicoid* mRNA, we have chosen the value, $D = 10^{-11}$ m^{2}s^{-1}, as estimated below in (14). By (7), $d = a_2^2 D/L^2$, and the value of the degradation rate d depends on the choices made for D and L.

As the experimental data is given in arbitrary light intensity units, the initial value of the *bicoid* mRNA concentration is also arbitrary. However. it is plausible to assume that the total amount of initial *bicoid* mRNA deposited in the embryo does not change too much for different embryos. So, in order to estimate the total amount of *bicoid* mRNA in the embryos, using the first relation in (7), we have calculated the quantity,

$$Aa(\ell_2 - \ell_1)/L = a_1 d(\ell_2 - \ell_1)/L, \tag{10}$$

where $A(\ell_2 - \ell_1)$ is the total amount of initial *bicoid* mRNA deposited in the embryo, and the rate a should not change too much for different embryos. Therefore, if the quantity in (10), does not change too much for different data sets, it is an indication of the ability of the model to describe data sets with different phenotypes. In fact, as shown in Table 1, the quantity (10) is almost constant among embryos, even if $(\ell_2 - \ell_1)/L$ shows a large variability, as is the case of the fits in Figure 2.

From the fitted values shown in Table 1, each of the calculated parameter values $Aa(\ell_2 - \ell_1)/L$, a_2 and d have the same order of magnitude for the different cleavage stages. In fact, the parameters defined in the kinetic mechanisms (2) are independent of the cleavage stage and, therefore, their values must depend only of the phenotypic characteristics of the analyzed embryo. The similarities between the parameters $Aa(\ell_2 - \ell_1)/L$, a_2 and d for different cleavage stages is an indication of the consistency of the theoretical model proposed here.

Determination of the diffusion coefficient of *bicoid* mRNA

To determine the value of the diffusion coefficient of the *bicoid* mRNA, we use the information that Bicoid protein reaches a steady state in approximately T seconds. So, we integrate the two equations in (1) along the embryo length, and using the zero flux boundary conditions, we obtain,

$$\begin{cases} \dfrac{d\bar{R}}{dt} &= -d\bar{R} \\ \dfrac{d\bar{B}}{dt} &= a\bar{R} \end{cases} \tag{11}$$

where $\bar{R}$ and $\bar{B}$ are the total amount of mRNA and protein in the embryo, respectively. The differential equations (11) have the solutions,

$$\begin{cases} \bar{R}(t) &= \bar{R}(0)e^{-dt} \\ \bar{B}(t) &= \dfrac{a}{d}\bar{R}(0)\left(1-e^{-dt}\right) \end{cases} \tag{12}$$

Assuming that the steady state of the Bicoid protein is attained after T seconds of development, and that 95% of the mRNA has been translated into protein, by (12), we have the development time relation, $\bar{B}(T)/(a\bar{R}(0)/d) = 0.95 = \left(1-e^{-dT}\right)$. From the previous relation we obtain, $d = -\log(0.05)/T$. Therefore, by (7), the diffusion coefficient is,

$$D = d\frac{L^2}{a_2^2} = -\frac{\log(0.05)}{T}\frac{L^2}{a_2^2} \tag{13}$$

As Bicoid protein attains the steady state at the end of cleavage stage 9, in approximately $T \simeq 90 \times 60$ seconds, (Gregor *et al.*, 2007), with the data in Table 1, we have $a_2 \in [3, 5.5]$, and with the choice $L = 0.5 \times 10^{-3}$ m, by (13), the diffusion coefficient of *bicoid* mRNA is in the range,

$$D \in [4.6 \times 10^{-12}, 1.5 \times 10^{-11}] \tag{14}$$

These estimates, as well as the numerical fits of Figure 2, are consistent with the theoretical predictions for the order of magnitude of the diffusion coefficients of large molecules (*bicoid* mRNA) in the cytoplasm (Wojcieszyn *et al.*, 1981; Mastro *et al.*, 1984).

Discussion

We have proposed a new model describing the production and the establishment of the stable gradient of the Bicoid protein along the antero-posterior axis of the embryo of *Drosophila*. In this model, *bicoid* mRNA diffuses along the antero-posterior axis of the embryo and Bicoid protein is produced and stays localized near the syncytial nuclei as observed in experiments.

We have calculated the steady state of the Bicoid protein along the antero-posterior axis of the embryo of *Drosophila*, and we have calibrated the parameters of the mRNA diffusion model with experimental data taken during cleavage stages 11-14. After the calibration of the model with experimental data, we have predicted the initial localization in the embryo of the *bicoid* mRNA of maternal origin (parameters ℓ_1 and ℓ_2 in Table 1), the Bicoid protein concentration profiles along the embryo, and the *bicoid* mRNA degradation rates. The mean relative errors between the theoretical prediction of the Bicoid protein steady state and the experimental data (Figure 2) are in the range $5\% - 8\%$, suggesting the effective validation of the model proposed here.

A simple estimate gives a diffusion coefficients D for *bicoid* mRNA in the interval $[4.6 \times 10^{-12}, 1.5 \times 10^{-11}]$ $\mathrm{m^2 s^{-1}}$. This estimate is calculated with the parameters found

in the calibration of the experimental data, with the model prediction formulas for the steady state, and with the additional assumption that the gradient of Bicoid protein is reached at the end of cleavage stage 9, (Gregor *et al.*, 2007). The determination of the diffusion coefficient is strongly dependent of the time duration of the cleavage cycles and therefore, it has a large error that is difficult to quantify. On the other hand, as the steady state solution of this *bicoid* mRNA diffusion model depends on the scaling parameter $a_2 = \sqrt{dL^2/D}$ and a_2 is determined by fitting the steady states of the protein profiles, the experimental determination of the degradation rate of *bicoid* mRNA leads to a more precise estimate of the diffusion coefficient.

The calibration and validation of the mRNA diffusion model shows that the mechanism of establishment of the gradient of Bicoid protein observed in *Drosophila* early development can be justified by a diffusion hypothesis for mRNAs. The mathematical model considers that *bicoid* mRNA diffuses along the embryo and the translated protein stays localized near the syncytial nuclei, as observed in Figure 2. In this model, protein degradation is not considered and proteins do not diffuse along the embryo. The model proposed here explains the experimental data for protein gradients, and the common assumption that morphogen gradients are obtained with a balanced and continuous production and degradation of proteins is not necessary. The low level of Bicoid concentration in the intranuclear regions of the embryo is easily explained through the ribosome localization near the syncytial nuclei.

Random motion of *bicoid* mRNA has been observed, (Cha *et al.*, 2001; Saxton, 2001), and the *bicoid* mRNA gradient has been recently found by Spirov *et al.* (Spirov *et al.*, 2009). The mechanism of mRNA diffusion proposed here is uni-dimensional and together with the very good agreement between the experimental data and the model predictions, we can raise the hypothesis of the existence of a mechanism of constrained mRNA diffusion along a network of nonpolar microtubules. This hypothesis has been discussed by Spirov *et al.* (Spirov *et al.*, 2009) and is consistent with a mechanism based on mRNA diffusion along microtubules. This justifies the very good agreement between one-dimensional diffusion models and the observed experimental data (the motion of mRNA is observed along the embryo wall). Two and three-dimensional diffusion models would predict protein concentrations in the interior of the embryo, which is not observed.

Materials and Methods

To fit the sets of data points of Figure 2, we consider that each data set is approximated by a function $B(x;\vec{\alpha})$, with $x \in [0,1]$ and where $\vec{\alpha} = (\alpha_1, \ldots, \alpha_m)$ is the set of m parameters to be determined. We assume that the parameter space $S = \{\vec{\alpha} : \infty < m_i \leq \alpha_i \leq M_i < \infty, i = 1, \ldots m\}$ is a compact subset of R^m. For each fixed value of the vector parameter $\vec{\alpha}$, and experimental data points $\{(x_i, B_{exp}(x_i))\}_{i=1}^n$, we consider the fitness function (sum

of the mean squared deviations),

$$\chi^2(\vec{\alpha}) = \frac{1}{n}\sum_{i=1}^{n}(B(x_i;\vec{\alpha}) - B_{exp}(x_i))^2 \tag{15}$$

The set of parameter values that best fits the experimental data is determined from the global minimization condition,

$$\min_{\vec{\alpha}\in S}\chi^2(\vec{\alpha})$$

In order to search the global minimum of the function $\chi^2(\vec{\alpha})$, with $\vec{\alpha} \in S$, we take a set of p vectors $\vec{\alpha}_k$, with $k = 1, \ldots, p$, randomly equidistributed in the set S. Then, for each $\vec{\alpha}_k$, we compute the fitness function (15).

To search for the global minimum of the mean squared deviation $\chi^2(\vec{\alpha})$, we swarm the set of p vectors $\vec{\alpha}_k$ in the parameter space S. For each vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_m)$, we construct a new vector $\vec{\alpha}'$ according to the swarm rule,

$$\alpha_i' = \alpha_i + \Delta t(M_i - m_i)(2\xi - 1) \tag{16}$$

where Δt is a time parameter, $(M_i - m_i)$ is a scaling constant, and ξ is a random variable uniformly distributed in the interval $[0, 1]$. Then, we recalculate the new value of the fitness function, $\chi^2(\vec{\alpha}')$. If $\chi^2(\vec{\alpha}') < \chi^2(\vec{\alpha})$, the parameter value $\vec{\alpha}$ is updated to the new value $\vec{\alpha}'$. If $\chi^2(\vec{\alpha}') \geq \chi^2(\vec{\alpha})$, no update is done. We repeat this procedure for all the parameter values in the search space S.

After iterating the swarm algorithm M times for all the population of parameter values, we order the parameter vectors according to their fitness values, and we discard half of the parameters that have the worst fitnesses. We repeat this procedure s times. The parameter values that best fit the experimental data are the ones that corresponds to the minimum of $\chi^2(\vec{\alpha})$.

This simple algorithm relies on the assumption that the initial number of random points are equidistributed in S and they form a sufficiently dense set in S.

In the cases in Figure 2, the convergence of the swarm algorithm for the determination of the global minimum of the fitness function (15) as a function of the parameters has been checked by graphical methods in two-dimensional sections of the parameter space S. In all the fits in Figure 2, the best convergence has been obtained with the swarm parameters, $p = 1024$, $M = 500$, $s = 5$ and $\Delta t = 0.01$, and good convergence for the global minima has been obtained.

A further extension of this calibration technique using evolutionary algorithms for the Bicoid-Caudal protein regulation in *Drosophila* has been developed in (Dilao *et al.*, 2009).

Acknowledgments

We thank Fred Cummings, Ana Pombo and Solveig Thorsteinsdottir for enlightening discussions during the preparation of this paper. This work has been supported by European

project GENNETEC, FP6 STREP IST 034952.

References

Alves, F. & Dilão, R. 2006 Modeling segmental patterning in *Drosophila*: Maternal and gap genes. *J. Theor. Biol.* **241**, 342-359.

Cha, B.-J., Koppetsch, B. S. & Theurkauf, W. E. 2001 In Vivo Analysis of *Drosophila bicoid* mRNA Localization Reveals a Novel Microtubule-Dependent Axis Specification Pathway. *Cell* **106**, 35-46.

Crick, F. 1970 Diffusion in embryogenesis. *Nature* **225**, 420–422.

Dilão, R., Muraro, D., Nicolau, M. & Schoenauer, M. 2009 Validation of a morphogenesis model of Drosophila early development by a multi-objective evolutionary optimization algorithm. In Pizzuti, C., Ritchie, M. D. & and Giacobini M. (Eds.): EvoBIO 2009, *Lecture Notes in Computer Science* **5483**, 176-190.

Driever, W. & Nüsslein-Volhard, C. 1988 A gradient of Bicoid protein in *Drosophila* embryos. *Cell* **54**, 83–93.

Forrest, K. M. & Gavis, E. R. 2003 Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for *nanos* mRNA localization in *Drosophila*. *Curr. Biol.* **13**, 1159-1168.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. & Noll, M. 1986 Structure of the segmentation gene paired and the Drosphila PRD gene set as part of a gene network. *Cell* **47**, 735–746.

Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W. & Tank, D. W. 2007 Stability and nuclear dynamics of the Bicoid morphogen gradient. *Cell* **130**, 141–152.

Haberman, R. 1998 *Elementary Applied Partial Differential Equations*. Upper Saddle River, NJ: Prentice-Hall.

Houchmandzadeh, B., Wieschaus, E. & Leibler, S. 2002 Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature* **415**, 798–802.

Houchmandzadeh, B., Wieschaus, E. & Leibler, S. 2005 Precise domain specification in the developing *Drosophila* embryo. *Phys. Rev. E* **72**, 061920.

Kerszberg, M. & Wolpert, L. 2007 Specifying Positional information in the embryo:Looking Beyond Morphogens. *Cell* **130**, 205–209.

Kozlov, K., Myasnikova, E., Samsonova, M., Reinitz, J., Kosman, D. 2000 Method for spatial registration of the expression patterns of Drosophila segmentation genes using wavelets. *Computational Technologies* **5**, 112–119.

Mastro, A. M., Babich, M. A., Taylor, W. D. & Keith, A. D. 1984 Diffusion of a small molecule in the cytoplasm of mammalian cells. *Proc. Natl. Acad. Sci. USA* **81**, 3414-3418.

Meinhardt, H. 1977 A model for pattern formation in insect embryogenesis. *J. Cell Sci.* **23**, 117–139.

Murray, J. D. 1993 *Mathematical Biology*. Berlin: Springer-Verlag.

Myasnikova, E., Kosman, D., Reinitz, J. & Samsonova, M. 1999 Spatio-temporal registration of the expression patterns of *Drosophila* segmentation genes. In Lengauer, T. & Schneider, R., Bork, P., Brutlag, D., Glasgow, J., Mewes, H.-W. & Zimmer, R. (eds), *Seventh International Conference on Intelligent Systems for Molecular Biology*, pp. 195–201. Menlo Park: AAAI Press.

Myasnikova, E., Samsonova, A., Kozlov, K., Samsonova, M. & Reinitz, J. 2001 Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods. *Bioinformatics* **17**(1), 3–12.

Nüsslein-Volhard, C. 1992 Gradients that organize embryo development. *Scientific American* **275**(2), 54–61.

2006 *Coming to life. How Genes Drive Development*. New Haven: Yale University Press, New Haven.

Pisarev, A., Poustelnikova, E., Samsonova, M. & Reinitz, J. 2009 FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. *Nucleic Acids Research* **37**, D560-D566.

Poustelnikova, E., Pisarev, A., Blagov, M., Samsonova, M. & Reinitz, J. 2004 A database for management of gene expression data in situ. *Bioinformatics* **20**, 2212–2221.

Reinitz, J. 2007 A ten per cent solution. *Nature* **448**, 420–421.

Saxton, W. M. 2001 Microtubules, Motors, and mRNA Localization Mechanisms: Watching Fluorescent Messages Move. *Cell* **107**, 707–710.

Spirov, A., Fahmy, K., Schneider, M., Frei, E., Nooll, M. & Baumgartner, S. 2009 Formation of the *bicoid* morphogen gradient: an mRNA gradient dictates the protein gradient. *Development* **136**, 605-614.

Turing, A. M. 1952 The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. London B* **237**, 37–72.

Wojcieszyn, J. W., Schlegel, R. A., Wu, E. S. & Jacobson, K. A. 1981 Diffusion of injected macromolecules within the cytoplasm of living cells. *Proc. Natl. Acad. Sci. USA* **78**, 4407–4410.

Wolpert, L. 1969 Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47.

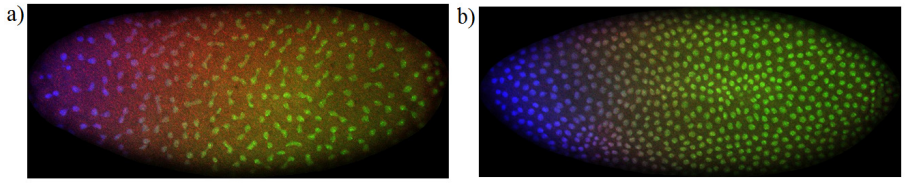

Figure 1: **Distribution of Bicoid protein (in blue) in the embryo of *Drosophila*, in the interphase following the cleavage stages 11 (a) and 12 (b).** The images are from the FlyEx datasets ab18 (a) and ab17 (b), (Kozlov *et al.*, 2000; Myasnikova *et al.*, 1999, 2001; Poustelnikova *et al.*, 2004; Pisarev *et al.*, 2009). Note the absence of Bicoid protein in the inter-nuclear regions of the cytoplasm. The localization of Bicoid protein near the nuclear envelopes suggest that ribosomes are also localized near the nucleus.

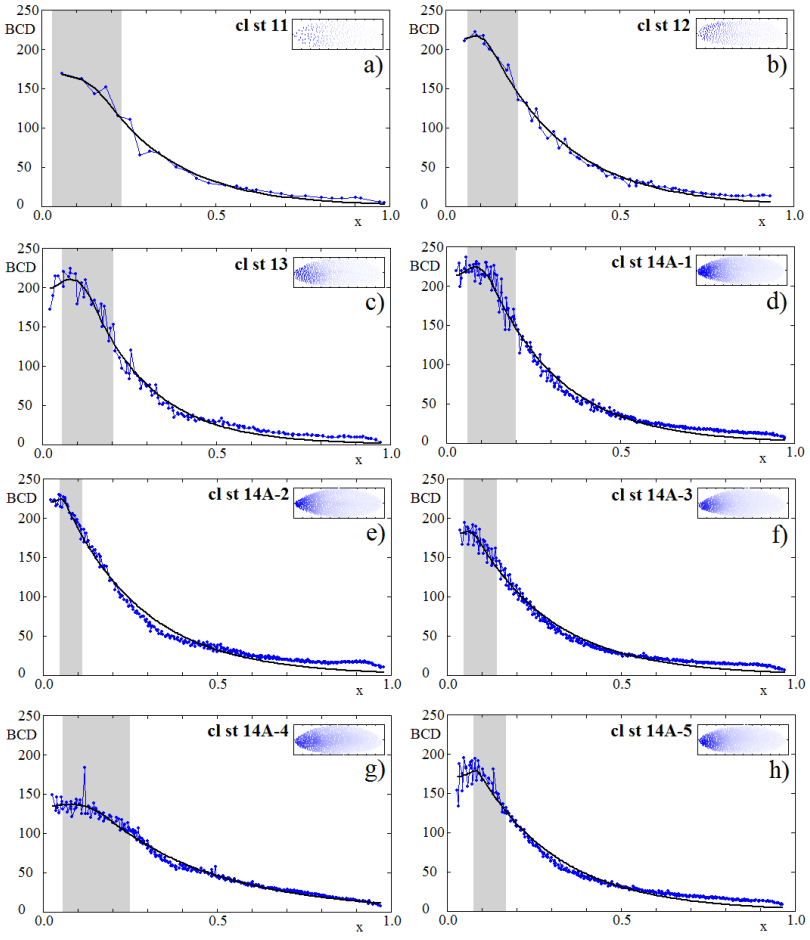

Figure 2: **Concentration of Bicoid protein (BCD) along the antero-posterior axis (x) of *Drosophila*, in the interphase following the consecutive cleavage stages (cl st) 11, 12, 13 and 14.** Data (connected dots) in the figures are from the FlyEx datasets: a) ab18 (11); b) ab17 (12); c) ab16 (13); d) ab12 (14A-1); e) ab14 (14A-2); f) ab9 (14A-3); g) ad13 (14A-4); h) ab8 (14A-5), and the numbers inside the parenthesis refer to the cleavage stage, (Kozlov *et al.*, 2000; Myasnikova *et al.*, 1999, 2001; Poustelnikova *et al.*, 2004; Pisarev *et al.*, 2009). In the upper right corner of the figures, we show the corresponding gradients of Bicoid protein in the two-dimensional projections of the embryo. The data points of the Bicoid gradients are taken from a region centered around the central antero-posterior axis of the embryo. The transversal length of this region is equal to 10% of the maximal length of the dorso-ventral direction. The lengths of the embryos have been rescaled to the value $L = 1$. The thick black lines are the best fits of the experimental data with the theoretical prediction (8)-(9). The gray regions show the initial localization of *bicoid* mRNA, and are defined by the fitted values ℓ_1/L and ℓ_2/L. The parameters of the fits are shown in Table 1.